

# NON-PARAMETRIC BAYESIAN DRIFT ESTIMATION FOR STOCHASTIC DIFFERENTIAL EQUATIONS

SHOTA GUGUSHVILI AND PETER SPREIJ

**ABSTRACT.** We consider non-parametric Bayesian estimation of the drift coefficient of a one-dimensional stochastic differential equation from discrete-time observations on the solution of this equation. Under suitable regularity conditions that are weaker than those previously suggested in the literature, we establish posterior consistency in this context. Furthermore, we show that posterior consistency extends to the multidimensional setting as well, which, to the best of our knowledge, is a new result in this setting.

## 1. INTRODUCTION

Consider the  $d$ -dimensional stochastic differential equation

$$(1) \quad dX_t = b(X_t)dt + dW_t$$

driven by a  $d$ -dimensional Brownian motion  $W$ , and assume that it has a unique (in the sense of the probability law) non-exploding weak solution. One can start with a coordinate mapping process  $X$  (that is  $X_t(\omega) = \omega(t)$ ) on the canonical space  $(\mathcal{C}(\mathbb{R}_+), \mathcal{B}(\mathcal{C}(\mathbb{R}_+)))$  of continuous functions  $\omega : \mathbb{R}_+ \rightarrow \mathbb{R}^d$ , a flow of sigma-fields  $\{\mathcal{F}_t^X\}$  and the  $d$ -dimensional Wiener measure  $Q$  on  $(\mathcal{C}(\mathbb{R}_+), \mathcal{B}(\mathcal{C}(\mathbb{R}_+)))$ , and then, as is well-known (see e.g. Proposition 3.6 and Remark 3.7 on p. 303 in Karatzas and Shreve (1988)), under suitable conditions on the drift coefficient  $b$  and for any fixed initial distribution  $\mu$  one can obtain a weak solution  $(X, W)$ ,  $(\mathcal{C}(\mathbb{R}_+), \mathcal{F}, P_b^\mu)$ ,  $\{\mathcal{F}_t\}$  to (1) through the Girsanov theorem. The filtration  $\{\mathcal{F}_t\}$  can be made to satisfy the usual conditions by suitably augmenting and completing the filtration  $\{\mathcal{F}_t^X\}$ , cf. Remark 3.7 on p. 303 in Karatzas and Shreve (1988). Henceforth we will assume that we are in this canonical setup. We will also assume that  $X$  is ergodic with a unique ergodic distribution  $\mu_b$  and is in fact initialised at  $\mu_b$ , so that  $\mu = \mu_b$ . Furthermore, we will abbreviate  $P_b^{\mu_b}$  to  $P_b$ .

Suppose that the drift coefficient  $b = (b_1, \dots, b_d)$  belongs to some non-parametric class. Denote by  $b_0 = (b_{0,1}, \dots, b_{0,d})$  the true drift coefficient and assume that corresponding to it a sample  $X_0, X_\Delta, X_{2\Delta}, \dots, X_{n\Delta}$  is given. The goal is to estimate  $b_0$  non-parametrically. The problem of non-parametric estimation of  $b_0$  from discrete-time observations has received considerable attention in the literature. For frequentist approaches to the problem see for instance Comte et al. (2007), Gobet et al. (2004) and Jacod (2000) in the one-dimensional case ( $d = 1$ ) and Dalalyan and Reiß (2007) and Schmisser (2013) in the general multidimensional case ( $d \geq 1$ ). However, a non-parametric Bayesian approach to estimation of  $b_0$  is

---

*Date:* February 26, 2013.

*2000 Mathematics Subject Classification.* Primary: 62G20, Secondary: 62M05.

*Key words and phrases.* Drift coefficient; Non-parametric Bayesian estimation; Posterior consistency; Stochastic differential equation.

also possible, see e.g. van der Meulen et al. (2012), van der Meulen and van Zanten (2013) and van Zanten (2012). In particular, under appropriate assumptions on the drift coefficient  $b$ , the weak solution to (1) will admit transition densities  $p_b(t, x, y)$ , and employing the Markov property, the likelihood corresponding to the observations  $X_{i\Delta}$ 's can be written as

$$(2) \quad \pi_b(X_0) \prod_{i=1}^n p_b(\Delta, X_{(i-1)\Delta}, X_{i\Delta}),$$

where  $\pi_b$  denotes a density of the distribution  $\mu_b$  of  $X_0$  (under our conditions  $\pi_b$  and  $p_b$  will be strictly positive and finite, see Sections 2 and 3 for details). A Bayesian would put a prior  $\Pi$  on the class of drift coefficients, say  $\mathcal{X}$ , and obtain a posterior measure of any measurable set  $B \subset \mathcal{X}$  through Bayes' formula

$$(3) \quad \Pi(B|X_0, \dots, X_{n\Delta}) = \frac{\int_B \pi_b(X_0) \prod_{i=1}^n p_b(\Delta, X_{(i-1)\Delta}, X_{i\Delta}) \Pi(db)}{\int_{\mathcal{X}} \pi_b(X_0) \prod_{i=1}^n p_b(\Delta, X_{(i-1)\Delta}, X_{i\Delta}) \Pi(db)}.$$

Here we tacitly assume suitable measurability of the integrands, so that the integrals in (3) are well-defined. In the Bayesian paradigm, posterior encapsulates all the information required for inferential purposes. Once posterior is available, one can proceed with computation of Bayes point estimates, credible sets and other quantities of interest in Bayesian statistics.

It has been argued convincingly in Diaconis and Freedman (1986) and elsewhere that a desirable property of a Bayes procedure is posterior consistency. In our context this will mean that for every neighbourhood (in a suitable topology)  $U_{b_0}$  of  $b_0$ ,

$$\Pi(U_{b_0}^c | X_0, \dots, X_{n\Delta}) \rightarrow 0, \quad P_{b_0}\text{-a.s.}$$

as  $n \rightarrow \infty$  (see Sections 2 and 3 for details). That is, roughly speaking, a consistent Bayesian procedure asymptotically puts posterior mass equal to one on every fixed neighbourhood of the true parameter: the posterior concentrates around the true parameter. In an infinite-dimensional setting, such as the one we are dealing with, posterior consistency is a subtle property that depends in an essential way on a specification of the prior, see e.g. Diaconis and Freedman (1986). Note also that the notion of posterior consistency depends on the topology on  $\mathcal{X}$ . Ideally one would like to establish posterior consistency in strong topologies. An implication of posterior consistency is that even though two Bayesians might start with two different priors, the role of the prior in their inferential conclusions will asymptotically, with the sample size growing indefinitely, wash out, and the two will eventually agree. Furthermore, posterior consistency also implies that the centre (in an appropriate sense) of the posterior distribution is a consistent (in the frequentist sense) estimator of the true parameter. For an introductory treatment of posterior consistency see Wasserman (1998).

In the context of discretely observed scalar diffusion processes given as solutions to stochastic differential equations, posterior consistency has been recently addressed in van der Meulen and van Zanten (2013), while the case when a continuous record of observations from a scalar diffusion process is available was covered under various setups in van der Meulen et al. (2006), Panzar and van Zanten (2009) and Pokern et al. (2013), where in particular the contraction rates of the posterior were derived. The techniques used in the latter three papers are of little use in the case of discrete observations. The proof of posterior consistency in

van der Meulen and van Zanten (2013) is based on the use of martingale arguments in a fashion similar to Tang and Ghosal (2007), see also Ghosal and Tang (2006). The latter paper deals with posterior consistency for estimation of the transition density of an ergodic Markov process. The idea of using martingale arguments in the proofs of consistency of nonparametric Bayesian procedures goes back to Walker (2003) and Walker (2004) in the i.i.d. setting. On the other hand, a similarity between the arguments used in the proof of posterior consistency in Tang and Ghosal (2007) and van der Meulen and van Zanten (2013) is to a considerable extent on a conceptual level only: conditions for posterior consistency in Tang and Ghosal (2007) involve conditions on transition densities that typically cannot be transformed into conditions on the drift coefficients, because transition densities associated with stochastic differential equations are usually unknown in explicit form. Furthermore, in the setting of van der Meulen and van Zanten (2013), who deal with ergodic and strictly stationary scalar diffusion processes (in particular,  $X_0$  is initialised at the ergodic distribution of the process  $X$ ), one cannot assume that the density  $\pi_{b_0}$  of  $X_0$  is known (as done on p. 1714 in Tang and Ghosal (2007)), for that would completely determine the unknown drift coefficient  $b_0$ .

The assumption on the class of drift coefficients in Theorem 3.5 of van der Meulen and van Zanten (2013) (the latter deals with posterior consistency), namely uniform boundedness of the drift coefficients, is quite restrictive in that it excludes even such a prototypical example of a stochastic differential equation as the Langevin equation (here we assume  $d = 1$ )

$$(4) \quad dX_t = -\beta X_t dt + \sigma dW_t,$$

where  $\beta$  and  $\sigma$  are two constants. A solution to (4) is called an Ornstein-Uhlenbeck process, see Example 6.8 on p. 358 in Karatzas and Shreve (1988) and p. 397 there. Hence, there is room for improvement.

In this work we will show that under suitable conditions posterior consistency in the one-dimensional case still holds for the class of unbounded drift coefficients satisfying the linear growth condition. In particular, the case of the Langevin equation is covered. In our proof of posterior consistency we follow the same train of thought as initiated in Walker (2003) and Walker (2004), at the same time making use of ideas from Tang and Ghosal (2007) and especially from van der Meulen and van Zanten (2013). According to van der Meulen and van Zanten (2013), p. 51, the boundedness condition on the drift coefficients cannot be avoided in their approach due to technical reasons. Our analysis and contribution to the literature, however, shows that given a willingness to assume some reasonable and classical conditions on the drift coefficients, the case of unbounded drift coefficients can also be covered via techniques similar to those in van der Meulen and van Zanten (2013). Perhaps more importantly, under some extra, but standard assumptions in non-parametric inference for multidimensional stochastic differential equations (cf. Dalalyan and Reiß (2007) and Schmisser (2013)), we show that our analysis in the one-dimensional case extends to the multidimensional setting as well. To the best of our knowledge, this is a new result in this context.

The rest of the paper is organised as follows: in the next section we state our main result in the one-dimensional case, while Section 3 deals with the general multidimensional case. In Section 4 we provide a brief discussion on the obtained results. The proofs of the results from Sections 2 and 3 are given in Section 5.

Finally, Appendices A and B contain several auxiliary statements used in Section 5 together with their proofs.

## 2. POSTERIOR CONSISTENCY: ONE-DIMENSIONAL CASE

In this section we consider the one-dimensional case ( $d = 1$ ). The class of drift coefficients we will be looking at will be a subset of the class  $\tilde{\mathcal{X}}(K)$  introduced below.

**Definition 1.** *The family  $\tilde{\mathcal{X}}(K)$  consists of Borel-measurable drift coefficients  $b : \mathbb{R} \rightarrow \mathbb{R}$  possessing the following two properties:*

(a) *for some constant  $K > 0$  and  $\forall b \in \tilde{\mathcal{X}}(K)$ , the linear growth condition*

$$|b(x)| \leq K(1 + |x|)$$

*is satisfied, and*

(b) *for each  $b \in \tilde{\mathcal{X}}(K)$  there exist two constants  $r_b > 0$  and  $M_b > 0$ , such that*

$$b(x) \operatorname{sgn}(x) \leq -r_b, \quad \forall |x| \geq M_b$$

*holds.*

*Remark 1.* Analogously to considering  $L_p$ -spaces instead of  $\mathcal{L}_p$ -spaces, we will identify two functions  $b_1$  and  $b_2$  in  $\tilde{\mathcal{X}}(K)$ , if  $b_1 = b_2$  Lebesgue a.e.  $\square$

*Remark 2.* The class  $\tilde{\mathcal{X}}(K)$  is such that the case of the Langevin equation (4) with  $\sigma = 1$  is covered for parameter  $\beta$  ranging in the interval  $(0, K]$ .  $\square$

The goal of the following proposition is to show that when  $b \in \tilde{\mathcal{X}}(K)$ , a unique non-exploding weak solution to (1) exists and has certain desirable properties. Although, strictly speaking, a weak solution is a triple  $(X, W), (\Omega, \mathcal{F}, P_b^\mu), \{\mathcal{F}_t\}$ , in order to avoid cumbersome formulations, in the sequel we will at times take a liberty to call  $X$  itself a weak solution.

**Proposition 1.** *For each  $b \in \tilde{\mathcal{X}}(K)$ , where  $\tilde{\mathcal{X}}(K)$  is defined in Definition 1,*

- (1) *a unique non-exploding weak solution to (1) exists,*
- (2) *the weak solution  $X$  to (1) is ergodic with unique ergodic distribution  $\mu_b$  admitting a density  $0 < \pi_b(x) < \infty, x \in \mathbb{R}$  with respect to the Lebesgue measure, and*
- (3) *transition probabilities  $P_b(t, x, \cdot)$  are absolutely continuous with respect to the Lebesgue measure with densities  $0 < p_b(t, x, y) < \infty, (t, x, y) \in (0, \infty) \times \mathbb{R} \times \mathbb{R}$ .*

*Remark 3.* The contents of Proposition 1 are standard, but perhaps not available at one place in the literature. The linear growth condition (a) in Definition 1 is a standard assumption to ensure existence of a unique non-exploding weak solution to (1), see e.g. Proposition 3.6 and Remark 3.7 on p. 303, Theorem 5.15 on p. 341 and Remark 5.19 on p. 342 in Karatzas and Shreve (1988). In fact this assumption allows one to construct a weak solution via the Girsanov theorem as mentioned in the beginning of Section 1. Property (b) in Definition 1 is a classical assumption ensuring existence of a unique ergodic distribution for  $X$ , see e.g. Assumption  $(H^*)$  on p. 548 in Florens-Zmirou (1989). Finally, the two properties in Definition 1 also yield a short proof of part (3) of Proposition 1. See Section 5 for more details.  $\square$

*Remark 4.* Apart of its use in the proof of Proposition 1, the linear growth condition (a) on the drift coefficients in Definition 1 is also used e.g. when establishing formula (19) in the proof of Lemma A.4 in Appendix A, and more generally in those instances where we invoke the Girsanov theorem. Furthermore, property (b) from Definition 1 ensures that for  $b \in \tilde{\mathcal{X}}(K)$  the density  $\pi_b$  of the ergodic distribution  $\mu_b$  of  $X$  decays exponentially fast at infinity, cf. formula (9). Hence  $\mu_b$  has moments of all orders and for any  $b_1, b_2 \in \tilde{\mathcal{X}}(K)$ , the Kullback-Leibler divergence  $K(\mu_{b_1}, \mu_{b_2}) = \int_{\mathbb{R}} \pi_{b_1}(x) \log(\pi_{b_1}(x)/\pi_{b_2}(x)) dx$  is finite. This comes in handy in the proof of Lemma A.1 in Appendix A.  $\square$

*Remark 5.* Positivity of  $\pi_b$  and  $p_b$  formally justifies rewriting the likelihood as in (2) and allows us to employ the likelihood ratio  $L_n(b)$  in the proof of Theorem 1.  $\square$

*Remark 6.* Measurability of the mapping  $b \mapsto p_b(t, x, y)$  is a subtle property essential in (3), but it is difficult to ascertain it in a general setting. Therefore we will simply tacitly assume that all the quantities in (3) (or in other formulae where we integrate with respect to the prior) are suitably measurable.  $\square$

Since the notion of posterior consistency depends on a topology on the class of drift coefficients under consideration, we have to introduce the latter first. We will base our topology on the transition operators  $P_{\Delta}^b$ . Transition operators associated with (1) and acting on the class of bounded measurable functions  $f : \mathbb{R} \rightarrow \mathbb{R}$  are defined by

$$P_t^b f(x) = \int_{\mathbb{R}} p_b(t, x, y) f(y) dy.$$

We want our topology to separate distinct drift coefficients, which can be thought of as an identifiability condition. At the same time we want the posterior measure to concentrate on arbitrarily small neighbourhoods of the true parameter  $b_0$ . Fortunately, this will be possible with our choice of topology, as it will have the required separation property.

As it often happens in practice, it will be convenient in our case to define a topology not by directly specifying the open sets, but rather by specifying a subbase  $\tilde{\mathcal{U}}$  (for a notion of a subbase see e.g. p. 37 in Dudley (2002)).

**Definition 2.** Let  $\nu$  be a finite Borel measure on  $\mathbb{R}$  that assigns strictly positive mass to every non-empty open subset of  $\mathbb{R}$ , and let  $\mathcal{C}_{bdd}(\mathbb{R})$  denote the class of all bounded continuous functions on  $\mathbb{R}$ . For fixed  $b \in \tilde{\mathcal{X}}(K)$ ,  $f \in \mathcal{C}_{bdd}(\mathbb{R})$  and  $\varepsilon > 0$ , define

$$U_{f,\varepsilon}^b = \{\tilde{b} \in \tilde{\mathcal{X}}(K) : \|P_{\Delta}^{\tilde{b}} f - P_{\Delta}^b f\|_{1,\nu} < \varepsilon\}.$$

Here  $\|\cdot\|_{1,\nu}$  denotes the  $L_1$ -norm with respect to the measure  $\nu$ .

The following definition specifies a topology on  $\tilde{\mathcal{X}}(K)$ .

**Definition 3.** The topology  $\tilde{\mathcal{T}}$  on  $\tilde{\mathcal{X}}(K)$  is determined by the requirement that the family

$$\tilde{\mathcal{U}} = \{U_{f,\varepsilon}^b : f \in \mathcal{C}_{bdd}(\mathbb{R}), \varepsilon > 0, b \in \tilde{\mathcal{X}}(K)\}$$

is a subbase for  $\tilde{\mathcal{T}}$ .

*Remark 7.* The topology in Definition 3 clearly depends on the choice of the measure  $\nu$ , but since  $\nu$  is assumed to be fixed beforehand and its specific choice is not of great importance for subsequent developments, it is not reflected in our notation.  $\square$

*Remark 8.* The fact that Definition 3 is a valid definition follows from a standard result in general topology, Theorem 2.2.6 in Dudley (2002).  $\square$

*Remark 9.* The topology in Definition 3 has already been employed in van der Meulen and van Zanten (2013), who in that respect follow Section 6 in Tang and Ghosal (2007). For a  $\mathcal{C}^2$ -function  $f$  and a small  $\Delta$ ,

$$P_t^b f(x) - P_t^{\tilde{b}} f(x) \approx \Delta(b(x) - \tilde{b}(x))f'(x),$$

cf. p. 50 in van der Meulen and van Zanten (2013). Hence for a small  $\Delta$ , the topology  $\tilde{\mathcal{T}}$  in some sense resembles the topology induced by the  $L_1(\nu)$ -norm on  $\tilde{\mathcal{X}}(K)$ .  $\square$

In Lemma 1 given below we will show that the topology of Definition 3 has the Hausdorff property. This is perfectly sufficient for our purposes. For a notion of a Hausdorff space see e.g. p. 30 in Dudley (2002).

**Lemma 1.** *The topological space  $(\tilde{\mathcal{X}}(K), \tilde{\mathcal{T}})$  with  $\tilde{\mathcal{T}}$  as in Definition 3 is a Hausdorff space.*

We are ready to give the definition of posterior consistency used in the present work. For a definition of a neighbourhood used in it, see p. 26 in Dudley (2002).

**Definition 4.** *Let the prior  $\Pi$  be defined on a set  $\mathcal{X}(K) \subset \tilde{\mathcal{X}}(K)$  and let  $b_0 \in \mathcal{X}(K)$ . We say that posterior consistency holds at  $b_0$ , if for every neighbourhood  $U_{b_0}$  of  $b_0$  in the relative topology  $\mathcal{T} = \{A \cap \mathcal{X}(K) : A \in \tilde{\mathcal{T}}\}$  (with  $\tilde{\mathcal{T}}$  as in Definition 3) we have*

$$\Pi(U_{b_0}^c | X_0, \dots, X_{n\Delta}) \rightarrow 0, \quad P_{b_0}\text{-a.s.}$$

as  $n \rightarrow \infty$ .

We need yet another definition (the definition of a uniformly equicontinuous family of functions appearing in it can be found on p. 51 in Dudley (2002)).

**Definition 5.** *A family  $\mathfrak{F}$  of functions  $f : \mathbb{R} \rightarrow \mathbb{R}$  is called locally uniformly equicontinuous, if for any compact set  $F \subset \mathbb{R}$ , the restrictions  $f|_F$  of the functions  $f \in \mathfrak{F}$  to  $F$  form a uniformly equicontinuous family of functions; i.e., for every  $\varepsilon > 0$ , there exists a  $\delta > 0$ , such that the inequality*

$$\sup_{f \in \mathfrak{F}} \sup_{\substack{x, y \in F \\ |x - y| < \delta}} |f(x) - f(y)| < \varepsilon$$

holds.

The following will be a collection of drift coefficients we will be looking at in our first main result, Theorem 1.

**Definition 6.** *Let  $\mathcal{X}(K)$  be the collection of drift coefficients, such that  $\mathcal{X}(K) \subset \tilde{\mathcal{X}}(K)$  and  $\mathcal{X}(K)$  is a locally uniformly equicontinuous family of functions.*

*Remark 10.* Functions  $f$  belonging to some locally uniformly equicontinuous family  $\mathfrak{F}$  of functions are obviously continuous. If the family  $\mathfrak{F}$  is such that for every compact set  $F \subset \mathbb{R}$  the restrictions  $f|_F$  of the functions  $f \in \mathfrak{F}$  to  $F$  uniformly satisfy a Hölder condition (i.e. a Hölder condition with the same Hölder constants), then  $\mathfrak{F}$  is a locally uniformly equicontinuous family of functions.  $\square$

We summarise our assumptions.

**Assumption 1.** *Assume that*

- (a) *a unique in law non-exploding weak solution to (1) corresponding to each  $b \in \mathcal{X}(K)$  is initialised at the ergodic distribution  $\mu_b$ ,*
- (b)  *$b_0 \in \mathcal{X}(K)$  denotes the true drift coefficient,*
- (c) *a discrete-time sample  $X_0, \dots, X_{n\Delta}$  from the solution to (1) corresponding to  $b_0$  is available (we assume that we are in the canonical setup as in Section 1), and finally,  $\Delta$  is fixed and independent of  $n$ .*

The following is our first main result.

**Theorem 1.** *Let Assumption 1 hold and suppose that the prior  $\Pi$  on  $\mathcal{X}(K)$  is such that*

$$(5) \quad \Pi(b \in \mathcal{X}(K) : \|b - b_0\|_{2, \mu_{b_0}} < \varepsilon) > 0, \quad \forall \varepsilon > 0.$$

*Here  $\|\cdot\|_{2, \mu_{b_0}}$  denotes the  $L_2$ -norm with respect to measure  $\mu_{b_0}$ . Then posterior consistency as in Definition 4 holds.*

*Remark 11.* The fact that the members  $b$  of the parameter set  $\mathcal{X}(K)$  must satisfy the linear growth assumption for a uniform constant  $K$ , as well as the fact that  $\mathcal{X}(K)$  must be a locally uniformly equicontinuous family of functions, is unfortunate, as this excludes many interesting and popular priors in non-parametric Bayesian statistics (for instance the Gaussian process priors; cf. Panzar and van Zanten (2009)), but cannot be avoided with the current method of proof (cf. the remarks on p. 51 and p. 60 in van der Meulen and van Zanten (2013)). In fact, already in the parametric setting stronger conditions are used to theoretically justify validity of Bayesian computational approaches, such as the ones in Beskos et al. (2006) and Eraker (2001). We also remark that in the parametric estimation case from discrete-time observations, some domination conditions on the drift coefficients are still imposed in the asymptotic studies in the frequentist literature, that do not appear to be easily dispensable, except perhaps in simple cases like that of the Langevin equation (4); see e.g. Dacunha-Castelle and Florens-Zmirou (1986) and Florens-Zmirou (1989).  $\square$

*Remark 12.* Condition (5) on the prior  $\Pi$  is formulated in terms of the  $L_2(\mu_{b_0})$ -neighbourhoods, while the posterior consistency assertion returned by Theorem 1 is for the weak topology  $\mathcal{T}$ . However, by Remark 9, for a small  $\Delta$  the ‘discrepancy’ is not as dramatic as it may seem at the first sight.  $\square$

*Remark 13.* Since  $b_0$  is unknown, the prior  $\Pi$  must verify (5) at all parameter values  $b \in \mathcal{X}$ . Such priors do exist: since conditions of Theorem 1 are implied by conditions in Theorem 3.5 in van der Meulen and van Zanten (2013) (an assumption on the drift coefficients  $b$  ensuring ergodicity of  $X$  is not made explicit in van der Meulen and van Zanten (2013), see p. 47 there, but this does not cause any problems in our setting), two concrete examples of the prior  $\Pi$  can be found in Section 4 in van der Meulen and van Zanten (2013).  $\square$

*Remark 14.* The conditions of Theorem 1 cover the case of the Langevin equation (4) with  $\sigma = 1$  known and  $\beta$  an unknown parameter of interest.  $\square$

### 3. POSTERIOR CONSISTENCY: MULTIDIMENSIONAL CASE

In this section we turn our attention from the one-dimensional case to the general multidimensional case ( $d \geq 1$ ). The developments in this section are parallel to those in Section 2 and involve some repetitions, so we try to be relatively brief.

Our parameter set will be a subset of the class  $\tilde{\mathcal{X}}(K_1, K_2)$  of drift coefficients introduced below.

**Definition 7.** *The family  $\tilde{\mathcal{X}}(K_1, K_2)$  consists of Borel-measurable drift coefficients  $b : \mathbb{R}^d \rightarrow \mathbb{R}^d$  possessing the following three properties:*

(a) *for any  $b \in \tilde{\mathcal{X}}(K_1, K_2)$ , there exists a  $\mathcal{C}^3$ -function  $V_b : \mathbb{R}^d \rightarrow \mathbb{R}$ , such that*

$$C_b = \int_{\mathbb{R}^d} e^{-2V_b(u)} du < \infty,$$

*$|V_b(x)|$  grows not faster than a polynomial of  $\|x\|$  at infinity and  $b = -[\nabla V_b]^{tr}$ , where  $\nabla V_b$  is the gradient of  $V_b$  and  $tr$  denotes transposition;*

(b) *for any  $b \in \tilde{\mathcal{X}}(K_1, K_2)$ , there exist three constants  $r_b > 0$ ,  $M_b > 0$  and  $\alpha_b \geq 1$ , such that*

$$b(x) \cdot x \leq -r_b \|x\|^{\alpha_b}, \quad \forall \|x\| \geq M_b,$$

*where by dot we denote the usual scalar product on  $\mathbb{R}^d$  and  $\|x\|$  is the  $L_2$ -norm of a vector  $x \in \mathbb{R}^d$ ;*

(c) *there exist two constants  $K_1 > 0$  and  $K_2 > 0$ , such that for any  $b \in \tilde{\mathcal{X}}(K_1, K_2)$ ,*

$$\|b(x)\| \leq K_1(1 + \|x\|), \quad \left| \frac{\partial}{\partial x_j} b_i(x) \right| \leq K_2, \quad \forall x \in \mathbb{R}^d, \quad i, j = 1, \dots, d.$$

*Remark 15.* Assumptions made in Definition 7 are more than enough to guarantee existence of the unique (in the sense of the probability law) solution to (1). By Proposition 1 in Schmisser (2013), cf. p. 27 in Dalalyan and Reiß (2007), these assumptions also imply existence of the unique ergodic distribution  $\mu_b$  that has the density

$$\pi_b(x) = \frac{1}{C_b} e^{-2V(x)} > 0$$

with respect to the  $d$ -dimensional Lebesgue measure. In models in physics the function  $V_b$  has the interpretation of the potential energy of the system. Furthermore, Proposition 1.2 in Gobet (2002) implies existence of strictly positive transition densities  $p_b(t, x, y)$  associated with (1). Finally, for any  $b, \tilde{b} \in \mathcal{X}(K_1, K_2)$  we also have that the Kullback-Leibler divergence  $K(\mu_b, \mu_{\tilde{b}})$  is finite, which we use in the proof of Lemma B.4 in Appendix B.  $\square$

*Remark 16.* Compared to the case  $d = 1$  in Section 2, assumptions made in Definition 7 on the class of drift coefficients, when specified to the case  $d = 1$ , are somewhat stronger.  $\square$

*Remark 17.* Examples of multidimensional stochastic differential equations satisfying assumptions in Definition 7 are given in Section 5.2 in Schmisser (2013).  $\square$



Define the transition operators associated with (1) and acting on the class of bounded measurable functions  $f : \mathbb{R}^d \rightarrow \mathbb{R}$  by

$$P_t^b f(x) = \int_{\mathbb{R}^d} p_b(t, x, y) f(y) dy,$$

and for every fixed  $f \in \mathcal{C}_{bdd}(\mathbb{R}^d)$ ,  $\varepsilon > 0$ ,  $b \in \tilde{\mathcal{X}}(K_1, K_2)$  let

$$U_{f,\varepsilon}^b = \{\tilde{b} \in \tilde{\mathcal{X}}(K_1, K_2) : \|P_{\Delta}^{\tilde{b}} f - P_{\Delta}^b f\|_{1,\nu} < \varepsilon\},$$

where  $\nu$  is a fixed finite Borel measure on  $\mathbb{R}^d$  that assigns strictly positive mass to every non-empty open subset of  $\mathbb{R}^d$ . Define the topology  $\tilde{\mathcal{T}}$  on  $\tilde{\mathcal{X}}(K_1, K_2)$  through a subbase

$$\tilde{\mathcal{U}} = \{U_{f,\varepsilon}^b : f \in \mathcal{C}_{bdd}(\mathbb{R}^d), \varepsilon > 0, b \in \tilde{\mathcal{X}}(K_1, K_2)\}.$$

In complete analogy to Lemma 1, we have the following result.

**Lemma 2.** *The topological space  $(\tilde{\mathcal{X}}(K_1, K_2), \tilde{\mathcal{T}})$  with  $\tilde{\mathcal{X}}(K_1, K_2)$  as in Assumption 7 is a Hausdorff space.*

Let  $\mathcal{X}(K_1, K_2) \subseteq \tilde{\mathcal{X}}(K_1, K_2)$ , with the interpretation that  $\mathcal{X}(K_1, K_2)$  is our parameter set, and let  $\mathcal{T} = \{A \cap \mathcal{X}(K_1, K_2) : A \in \tilde{\mathcal{T}}\}$  be the corresponding relative topology on  $\mathcal{X}(K_1, K_2)$ . If for any neighbourhood  $U_{b_0} \in \mathcal{T}$  of  $b_0 \in \mathcal{X}(K_1, K_2)$  we have

$$\Pi(U_{b_0}^c | X_0, \dots, X_{n\Delta}) \rightarrow 0, \quad P_{b_0}\text{-a.s.}$$

as  $n \rightarrow \infty$ , we will say that posterior consistency holds at  $b_0$ .

We summarise our assumptions.

**Assumption 2.** *Assume that*

- (a) *a unique in law non-exploding weak solution to (1) corresponding to each  $b \in \mathcal{X}(K_1, K_2)$  is initialised at the ergodic distribution  $\mu_b$ ,*
- (b)  *$b_0 \in \mathcal{X}(K_1, K_2)$  denotes the true drift coefficient,*
- (c) *a discrete-time sample  $X_0, \dots, X_{n\Delta}$  from the solution to (1) corresponding to  $b_0$  is available (we assume that we are in the canonical setup as in Section 1), and finally,  $\Delta$  is fixed and independent of  $n$ .*

Under Assumption 2, the following multidimensional analogue of Theorem 1 holds.

**Theorem 2.** *Let Assumption 2 hold and suppose that the prior  $\Pi$  on  $\mathcal{X}(K_1, K_2)$  is such that*

$$(6) \quad \Pi \left( b \in \mathcal{X}(K_1, K_2) : \left\{ \sum_{i=1}^d \|b_i - b_{0,i}\|_{2,\mu_{b_0}}^2 \right\}^{1/2} < \varepsilon \right) > 0, \quad \forall \varepsilon > 0.$$

*Then posterior consistency holds.*

Condition (6) on the prior is of the same type as condition (5) in Theorem 1. We provide an example of a prior  $\Pi$  satisfying this condition. The construction of  $\Pi$  is similar to that in Example 4.1 in van der Meulen and van Zanten (2013). Both examples are related to discrete net priors in non-parametric Bayesian inference problems studied in Ghosal et al. (1997). The construction is admittedly artificial, but its sole goal is to show existence of a prior satisfying (6).

*Example 1.* Let  $\mathfrak{F}$  be a collection of  $\mathcal{C}^3$ -functions  $f : \mathbb{R} \rightarrow \mathbb{R}$ , such that

(a) for some polynomial function  $G : \mathbb{R} \rightarrow \mathbb{R}$  and  $\forall f \in \mathfrak{F}$  we have

$$|f(x)| \leq G(x), \quad \forall x \in \mathbb{R}_+;$$

(b) for some constant  $K_1 > 0$  and  $\forall f \in \mathfrak{F}$  we have

$$|f'(x)| \leq \frac{K_1}{2}, \quad \forall x \in \mathbb{R}_+;$$

(c)  $\forall f \in \mathfrak{F}$  we have

$$\int_{\mathbb{R}^d} e^{-2f(\|x\|^2)} dx < \infty;$$

(d)  $\forall f \in \mathfrak{F}$  there exist two constants  $M_f > 0$  and  $r_f > 0$ , such that  $f'(x) \geq r_f, \forall x \geq M_f$ ;

(e) for some constant  $K_2 > 0$  and  $\forall f \in \mathfrak{F}$ ,

$$\sup_{x \in \mathbb{R}_+} \{4x|f''(x)| + 2|f'(x)|\} \leq K_2.$$

For all  $x \in \mathbb{R}^d$  set  $V_f(x) = f(\|x\|^2)$  and  $b_f(x) = -[\nabla V_f(x)]^{tr}$ . Let  $\mathcal{X}(K_1, K_2)$  be a subset of a collection of all functions  $b_f = (b_{f,1}, \dots, b_{f,d})$  obtained in this way (the fact that this is a valid definition, in the sense that the requirements from Definition 7 are satisfied, follows by easy, but somewhat tedious computations; note that by taking  $f_\beta = \beta x^2/2$  and assuming  $d = 1$  and  $\beta \in (0, K_1]$ , we can cover the case of the Langevin equation (4)). We get from (b) that for every fixed  $i = 1, \dots, d$ , the functions  $b_{f,i}$  are locally bounded by constants uniform in  $f \in \mathfrak{F}$ . Furthermore, they are Lipschitz with uniform constants in  $f \in \mathfrak{F}$  as well: by the mean value theorem,

$$\begin{aligned} |b_{f,i}(x) - b_{f,i}(y)| &\leq \|\nabla b_{f,i}(\lambda x + (1-\lambda)y)\| \|x - y\| \\ &\leq \sqrt{d} K_2 \|x - y\|. \end{aligned}$$

Hence for each  $m \in \mathbb{N}$  and  $i = 1, \dots, d$ , by the Arzelà-Ascoli theorem, see Theorem 2.4.7 in Dudley (2002), the collection  $\mathfrak{B}_{m,i}$  of restrictions  $b_{f,i}|_{[-m,m]}$  of the functions  $b_{f,i}, f \in \mathfrak{F}$  to the intervals  $[-m, m]$  is totally bounded for the supremum metric  $\|\cdot\|_\infty$  (for the required definitions see p. 45 and p. 52 in Dudley (2002)). Then so is the product  $\bigotimes_{i=1}^d \mathfrak{B}_{m,i}$  for the product metric

$$\|b_f\|_{d,m,\infty} = \max_{i=1,\dots,d} \|b_{f,i}|_{[-m,m]}\|_\infty,$$

as well as its subset consisting of elements

$$b_f|_{[-m,m]^d} = (b_{f,1}|_{[-m,m]}, \dots, b_{f,d}|_{[-m,m]}), \quad f \in \mathfrak{F}.$$

Take a sequence  $\epsilon_l \downarrow 0$ . For any  $l \in \mathbb{N}$ , there exists a finite subset  $\mathfrak{F}_{m,\epsilon_l} = \{f_n^{m,\epsilon_l}, n = 1, \dots, n_{m,l}\}$  such that for any  $f \in \mathfrak{F}$ ,  $\|b_f - b_{f_n^{m,\epsilon_l}}\|_{d,m,\infty} < \epsilon_l$  for some  $n = 1, \dots, n_{m,l}$ . Let  $\tilde{Q}_1$  and  $\tilde{Q}_2$  be two measures on  $\mathbb{N}$ , such that  $q_{j,i} = \tilde{Q}_i(j) > 0, i = 1, 2, j \in \mathbb{N}$ . The prior  $\Pi$  on  $\mathcal{X}(K_1, K_2)$  is defined by

$$\Pi = \sum_{m=1}^{\infty} \sum_{l=1}^{\infty} \sum_{n=1}^{n_{m,l}} \frac{q_{m,1} q_{l,2}}{n_{m,l}} \delta_{b_{f_n^{m,\epsilon_l}}},$$

where  $\delta_{b_{f_n^{m,\epsilon_l}}}$  is the Dirac measure at  $b_{f_n^{m,\epsilon_l}}$ . The fact that  $\Pi$  satisfies requirement (5) of Theorem 1 is the content of Lemma 3 in Section 5. Since  $\Pi$  assigns all its mass to a countable subset of  $\mathcal{X}(K_1, K_2)$ , measurability issues should not concern us when integrating with respect to  $\Pi$ .  $\square$

## 4. DISCUSSION

In this work we were able to demonstrate that posterior consistency for non-parametric Bayesian estimation of the drift coefficient of a stochastic differential equation holds not only for the class of uniformly bounded drift coefficients and in the one-dimensional setting, as shown previously in van der Meulen and van Zanten (2013), but also in the multidimensional setting for the class of drift coefficients satisfying a linear growth assumption. This considerably enlarges the scope of the main result in van der Meulen and van Zanten (2013). Interestingly, although derivation of the posterior consistency result in the one-dimensional case is quite involved in van der Meulen and van Zanten (2013) (cf. a remark on p. 60 in van der Meulen and van Zanten (2013)) and replacement of the uniform boundedness condition on the drift coefficient with the linear growth condition requires some technical prowess in the proofs, see in particular the proof of Lemma A.4 in Appendix A, generalisation to the multidimensional setting does not involve technicalities far different from those in the one-dimensional setting, provided one suitably restricts the non-parametric class of drift coefficients. In fact, conditions we impose in the multidimensional setting are analogous to those used in the frequentist literature, see Dalalyan and Reiß (2007) and Schmisser (2013), which is a comforting fact. On the other hand, posterior consistency results both in van der Meulen and van Zanten (2013) and in our work are established for a weak topology on the class of drift coefficients. This is a consequence of the fact that we rely on techniques from Walker (2004) in our proofs, which are better suited for proving posterior consistency in weak topologies. Consistency in stronger topologies could have been established and contraction rates of the posterior could have been derived from general results for posterior consistency in Markov chain models had we known existence of certain tests satisfying conditions as in formula (2.2) in Ghosal and van der Vaart (2007); see Theorem 5 there. Existence of such tests for Markov chain models has been demonstrated in Theorem 3 in Birgé (1983), but unfortunately, the conditions involved in this theorem, cf. also formula (4.1) in Ghosal and van der Vaart (2007), do not appear to hold in general for the stochastic differential equation models such as van der Meulen and van Zanten (2013) and we are considering. Hence establishing posterior consistency in a stronger topology and derivation of the posterior contraction rate for non-parametric Bayesian drift estimation is an interesting and difficult open problem. A recent paper Pokern et al. (2013) addresses the latter question for a one-dimensional stochastic differential equation with a periodic drift coefficient. However, this is done under an assumption that an entire sample path  $\{X_t : t \in [0, T]\}$  is observed over the time interval  $[0, T]$  with  $T \rightarrow \infty$ . Moreover, periodic drift coefficients are completely different from the drift coefficients considered in Section 2 of the present work and making use of the techniques from Pokern et al. (2013) is impossible in our setting. Neither are the techniques in van der Meulen et al. (2006) and Panzar and van Zanten (2009) of any significant help (these papers deal with continuously observed scalar diffusion processes). It should also be noted that in the frequentist setting too (with  $\Delta$  fixed) already in the one-dimensional setting study of convergence rates of non-parametric estimators of the drift and dispersion coefficients is a highly non-trivial task, see e.g. Gobet et al. (2004), where various simplifying assumptions have been made, such as the requirement that the diffusion process under consideration has a compact state space, say  $[0, 1]$ , and is reflecting at the boundary points. Nevertheless, some

progress in establishing posterior consistency in a stronger topology than in the present work might be possible in the setting where  $\Delta = \Delta_n \rightarrow 0$  in such a way that  $n\Delta_n \rightarrow \infty$  (the so-called high-frequency data setting).

Finally, we remark that issues associated with practical implementation of the non-parametric Bayesian approach to estimation of a drift coefficient are outside the scope of the present work. Although much remains to be done in this direction, preliminary studies, such as the ones in Papaspiliopoulos et al. (2012) and van der Meulen et al. (2012), see also the overview paper van Zanten (2012), indicate that a non-parametric Bayesian approach in this context is both feasible and leads to reasonable results.

## 5. PROOFS

*Proof of Proposition 1.* As already mentioned in Remark 3, property (a) in Definition 1 suffices to guarantee existence of a unique non-exploding weak solution to equation (1). This proves part (1) of the proposition.

We next prove part (2). Although the result is well-known, a detailed proof does not seem to be available in the literature. We provide it for the reader's convenience. Introduce the scale function

$$s_b(y) = \int_0^y \exp \left( -2 \int_0^z b(x) dx \right) dz.$$

To show existence of an ergodic distribution, it is enough to show its existence for a process  $\tilde{X}_t = s_b(X_t)$ , cf. p. 48 in Skorokhod (1987). This process satisfies the stochastic differential equation

$$d\tilde{X}_t = \tilde{\sigma}(\tilde{X}_t) dW_t,$$

where

$$\tilde{\sigma}(y) = s'_b(s_b^{-1}(y)),$$

see Lemma 9 on p. 47 in Skorokhod (1987). Here  $s_b^{-1}$  denotes an inverse of  $s_b$ . If we can show that

$$(7) \quad s_b(-\infty) = -\infty, \quad s_b(\infty) = \infty, \quad \int_{\mathbb{R}} \frac{1}{\tilde{\sigma}^2(y)} dy < \infty,$$

then Theorem 16 on p. 51 in Skorokhod (1987) will imply existence of a unique ergodic distribution for  $\tilde{X}$ , and hence for  $X$  too. However, under our assumptions checking these conditions is easy. Assuming for instance that the first two conditions in (7) have been verified (the arguments used in their verification are similar to those used in verification of the third one), we will check the last one. By a change of the integration variable  $x = s_b^{-1}(y)$ , we have

$$(8) \quad \begin{aligned} \int_{\mathbb{R}} \frac{1}{\tilde{\sigma}^2(y)} dy &= \int_{\mathbb{R}} \frac{1}{s'_b(x)} dx \\ &= \int_{\mathbb{R}} \exp \left( 2 \int_0^x b(y) dy \right) dx. \end{aligned}$$

Finiteness of the latter integral can be seen as follows: let  $x \geq M_b > 0$  and note that

$$\begin{aligned}
 \exp\left(2 \int_0^x b(y) dy\right) &= \exp\left(2 \int_0^{M_b} b(y) dy\right) \exp\left(2 \int_{M_b}^x b(y) dy\right) \\
 (9) \quad &\leq \exp\left(2K \int_0^{M_b} (1+y) dy\right) \exp\left(-2r_b \int_{M_b}^x dy\right) \\
 &= c_{b,K} \exp(-r_b |x|).
 \end{aligned}$$

The argument for negative  $x$  is similar and yields a similar inequality. The integral in (8) is then finite thanks to the exponential decay property as in (9), and existence of a unique ergodic distribution follows. By the same Theorem 16 on p. 51 in Skorokhod (1987), the density of the ergodic distribution of  $\tilde{X}$  is given by

$$\tilde{\pi}(x) = \frac{(\tilde{\sigma}(x))^{-2}}{\int_{\mathbb{R}} (\tilde{\sigma}(x))^{-2} dx}.$$

By the fact that  $s_b$  is strictly increasing and hence  $P(X_t \leq x) = P(s_b(X_t) \leq s_b(x))$ , it then follows by differentiation that the density of the ergodic distribution of  $X$  is given by

$$\pi_b(x) = \frac{1}{m_b(\mathbb{R})} \exp\left(2 \int_0^x b(y) dy\right),$$

where

$$m_b(dx) = \exp\left(2 \int_0^x b(y) dy\right) dx$$

is the speed measure of  $X$ . Furthermore, by the linear growth condition,  $0 < \pi_b(x) < \infty, \forall x \in \mathbb{R}$ . This proves part (2).

Finally, for the proof of part (3) we argue as follows: the first two equalities in (7) and Proposition 5.22 (a) on p. 345 in Karatzas and Shreve (1988) yield that the process  $X$  is recurrent. Hence the solution to (1) generates a regular diffusion (see Definition 45.2 on p. 272 in Rogers and Williams (1987)). By Theorem 50.11 on pp. 294–295 in Rogers and Williams (1987), the transition probabilities of  $X$  admit continuous, strictly positive and finite densities with respect to the speed measure  $m_b(dy)$  of  $X$ . Since from part (2) we have in turn that

$$0 < \exp\left(2 \int_0^y b(z) dz\right) < \infty, \quad \forall y \in \mathbb{R},$$

it follows that transition probabilities of  $X$  admit continuous, strictly positive and finite densities  $p_b(t, x, y)$  with respect to the Lebesgue measure. This completes the proof of the proposition.  $\square$

*Proof of Lemma 1.* The lemma can be proved by arguments similar to those in the proof of Lemma 3.2 in van der Meulen and van Zanten (2013); cf. also the proof of Lemma 2 in Appendix B. The following result, which is an analogue of Lemma 3.1 in van der Meulen and van Zanten (2013), is required in the proof (the arguments from the proof of the latter remain applicable): let  $b \in \tilde{\mathcal{X}}(K)$ . Fix  $t > 0$ . If  $b \neq \tilde{b}$ , then  $P_t^b \neq P_t^{\tilde{b}}$ .  $\square$

*Proof of Theorem 1.* The proof follows the same main steps as the proof of Theorem 3.5 in van der Meulen and van Zanten (2013), which in turn uses some ideas from Tang and Ghosal (2007) and Walker (2004). In particular, our proof employs Lemmas A.1, A.2 and A.3 from Appendix A, that correspond to Lemmas 5.1, 5.2 and 5.3 in van der Meulen and van Zanten (2013). Fix  $\varepsilon > 0$ , take a fixed  $f \in \mathcal{C}_{bdd}(\mathbb{R})$  and write

$$(10) \quad B = \{b \in \mathcal{X}(K) : \|P_{\Delta}^b f - P_{\Delta}^{b_0} f\|_{1,\nu} > \varepsilon\}.$$

Without loss of generality we may assume that  $\|f\|_{\infty} \leq 1$  and  $\varepsilon \leq 2\nu(\mathbb{R})$ . We claim that by the definition of the topology  $\mathcal{T}$  it suffices to establish posterior consistency for every fixed  $B$  of the above form. Indeed, by intersecting the sets from the base  $\tilde{\mathcal{V}}$  determined by the subbase  $\tilde{\mathcal{U}}$  from Definition 3 with  $\mathcal{X}(K)$ , a base for  $\mathcal{T}$  can be obtained. Likewise, a subbase  $\mathcal{U}$  for  $\mathcal{T}$  is obtained by intersecting the sets from the subbase  $\tilde{\mathcal{U}}$  for  $\tilde{\mathcal{T}}$  with  $\mathcal{X}(K)$ . By definition, an arbitrary neighbourhood  $U_{b_0}$  of  $b_0$  contains an open set  $\hat{U}_{b_0} \in \mathcal{T}$ . The set  $\hat{U}_{b_0}$  is a union of open sets  $V$  from the base  $\mathcal{V}$ ,  $\hat{U}_{b_0} = \bigcup\{V \in \mathcal{V} : V \subset \hat{U}_{b_0}\}$ . There is at least one  $V$  that contains  $b_0$ . Fix such  $V$ . By definition of the subbase  $\mathcal{U}$  this set  $V$  can be represented as  $V = \bigcap_{j=1}^m U_{f_j, \varepsilon_j}^{b_0}$  for some  $m$ , positive numbers  $\varepsilon_j$ , bounded continuous functions  $f_j$  and sets  $U_{f_j, \varepsilon_j}^{b_0}$  from the subbase  $\mathcal{U}$ . Note that we have

$$U_{b_0}^c \subset \hat{U}_{b_0}^c \subset V^c = \bigcup_{j=1}^m (U_{f_j, \varepsilon_j}^{b_0})^c.$$

Since

$$\begin{aligned} (U_{f_j, \varepsilon_j}^{b_0})^c &= \{b \in \mathcal{X}(K) : \|P_{\Delta}^b f_j - P_{\Delta}^{b_0} f_j\|_{1,\nu} \geq \varepsilon_j\} \\ &\subset \left\{b \in \mathcal{X}(K) : \|P_{\Delta}^b f_j - P_{\Delta}^{b_0} f_j\|_{1,\nu} > \frac{\varepsilon_j}{2}\right\}, \end{aligned}$$

say, the claim becomes obvious.

The posterior measure of a set  $B$  given in (10) can be written as

$$\Pi(B|X_0, \dots, X_{n\Delta}) = \frac{\int_B L_n(b) \Pi(db)}{\int_{\mathcal{X}(K)} L_n(b) \Pi(db)},$$

where

$$L_n(b) = \frac{\pi_b(X_0)}{\pi_{b_0}(X_0)} \prod_{i=1}^n \frac{p_b(\Delta, X_{(i-1)\Delta}, X_{i\Delta})}{p_{b_0}(\Delta, X_{(i-1)\Delta}, X_{i\Delta})}$$

is the likelihood ratio. By Lemma A.3 in Appendix A, in order to prove the theorem, it suffices to show that

$$\Pi(B_j^+|X_0, \dots, X_{n\Delta}) \rightarrow 0, \quad \Pi(B_j^-|X_0, \dots, X_{n\Delta}) \rightarrow 0, \quad P_{b_0}\text{-a.s.}$$

for the sets  $B_j^+$  and  $B_j^-$  ( $j = 1, \dots, N$  for some suitable integer  $N > 0$ ) given in the statement of that lemma. We give a brief outline of the remaining part of the proof: thanks to property (5) of the prior, by Lemma A.1 from Appendix A the prior  $\Pi$  has the Kullback-Leibler property in the sense that (11) holds. Then by Lemma A.2 from Appendix A in order to establish posterior consistency, it suffices to show that  $P_{b_0}$ -a.s. the terms

$$\sqrt{\int_{B_j^+} L_n(b) \Pi(db)}, \quad \sqrt{\int_{B_j^-} L_n(b) \Pi(db)},$$

converge to zero exponentially fast. This fact can be proved by the same reasoning as given in the proof of Theorem 3.5 in van der Meulen and van Zanten (2013) (employing the convergence theorem for a positive supermartingale (see e.g. Theorem 22 on p. 148 in Pollard (2002)) instead of Doob's martingale convergence theorem on pp. 59–60 there<sup>1</sup>). This completes the proof.  $\square$

*Proof of Lemma 2.* The lemma can be proved by arguments similar to those in the proof of Lemma 3.2 in van der Meulen and van Zanten (2013). The proof employs Lemma B.1 from Appendix B, that plays the role of Lemma 3.1 from van der Meulen and van Zanten (2013) in this context.  $\square$

*Proof of Theorem 2.* The proof is an easy generalisation of the proof of Theorem 1 and uses lemmas from Appendix B instead of Lemmas A.1, A.2 and A.3 from Appendix A.  $\square$

In the next lemma we verify the claim made at the end of Example 1.

**Lemma 3.** *The prior  $\Pi$  from Example 1 satisfies the requirement (6).*

*Proof.* The proof is similar to a demonstration of an analogous property of the prior in Example 4.1 in van der Meulen and van Zanten (2013): for every  $b \in \mathcal{X}(K_1, K_2)$  and positive integer  $m$  we have

$$\begin{aligned} \sum_{i=1}^d \|b_i - b_{i,0}\|_{2,\mu_{b_0}}^2 &= \sum_{i=1}^d \int_{\|x\| \leq m} (b_i(x) - b_{0,i}(x))^2 \pi_{b_0}(x) dx \\ &\quad + \sum_{i=1}^d \int_{\|x\| > m} (b_i(x) - b_{0,i}(x))^2 \pi_{b_0}(x) dx \\ &\leq d \|b - b_0\|_{m,d,\infty}^2 \\ &\quad + 4K^2 d \int_{\|x\| > m} (1 + \|x\|)^2 \pi_{b_0}(x) dx. \end{aligned}$$

Thanks to the fact that  $\mu_{b_0}$  has an exponential moment, the second term on the right-hand side can be made less than  $\varepsilon^2$  by choosing  $m$  large enough. Hence

$$\begin{aligned} \Pi \left( b \in \mathcal{X}(K_1, K_2) : \sum_{i=1}^d \|b_i - b_{0,i}\|_{2,\mu_{b_0}}^2 < 2\varepsilon^2 \right) \\ \geq \Pi \left( b \in \mathcal{X}(K_1, K_2) : \|b - b_0\|_{m,d,\infty}^2 < \frac{\varepsilon^2}{d} \right). \end{aligned}$$

For  $l$  such that  $\varepsilon_l < \varepsilon/\sqrt{d}$ , we have by construction of  $\Pi$  that the right-hand side of the above display is bounded from below by  $q_{m,1}q_{l,2}/k_{m,l} > 0$ . This completes the proof of the lemma.  $\square$

## APPENDIX A.

The following result is a restatement of Lemma 5.1 in van der Meulen and van Zanten (2013).

---

<sup>1</sup>Note that on p. 58 in van der Meulen and van Zanten (2013) the expression  $L_n$  is called the likelihood, although obviously the likelihood ratio is meant.

**Lemma A.1.** *Let*

$$\text{KL}(b_0, b) = \int_{\mathbb{R}} \int_{\mathbb{R}} \pi_{b_0}(x) p_{b_0}(\Delta, x, y) \log \frac{p_{b_0}(\Delta, x, y)}{p_b(\Delta, x, y)} dx dy.$$

*Then for the prior  $\Pi$  satisfying property (5), the inequality*

$$(11) \quad \Pi(b \in \mathcal{X}(K) : \text{KL}(b_0, b) < \varepsilon) > 0, \quad \forall \varepsilon > 0$$

*holds.*

*Proof.* The same proof as in van der Meulen and van Zanten (2013) goes through. The only additional clarification we would like to make concerns finiteness of  $\text{KL}(b_0, b)$ . The latter follows from the inequality in the proof of Lemma 5.1 in van der Meulen and van Zanten (2013),

$$\text{KL}(b_0, b) \leq -K(\mu_{b_0}, \mu_b) + K(\mathcal{L}_1, \mathcal{L}_2),$$

where

$$(12) \quad K(\mathcal{L}_1, \mathcal{L}_2) = \mathbb{E}_{P_{b_0}} \left[ \log \frac{\pi_{b_0}(X_0)}{\pi_b(X_0)} - \int_0^\Delta (b(X_s) - b_0(X_s)) dW_s + \frac{1}{2} \int_0^\Delta (b(X_s) - b_0(X_s))^2 ds \right] \\ = K(\mu_{b_0}, \mu_b) + \frac{\Delta}{2} \|b - b_0\|_{2, \mu_{b_0}}^2$$

is the Kullback-Leibler divergence between the laws  $\mathcal{L}_1 = \mathcal{L}(\{X_t, t \in [0, \Delta]\} | P_{b_0})$  and  $\mathcal{L}_2 = \mathcal{L}(X_t, t \in [0, \Delta] | P_b)$  of the full path  $\{X_t, t \in [0, \Delta]\}$  under  $P_{b_0}$  and  $P_b$ , respectively, while  $K(\mu_{b_0}, \mu_b)$  is the Kullback-Leibler divergence between the two invariant measures  $\mu_{b_0}$  and  $\mu_b$ . The second term on the right-hand side of the last equality in (12) is finite by the exponential decay property of  $\pi_{b_0}$ , cf. formula (9). Furthermore, we have

$$K(\mu_{b_0}, \mu_b) = \int_{\mathbb{R}} \pi_{b_0}(x) \log \frac{\pi_{b_0}(x)}{\pi_b(x)} dx \\ \leq \left| \log \frac{m_b(\mathbb{R})}{m_{b_0}(\mathbb{R})} \right| + 4K \int_{\mathbb{R}} \left( |x| + \frac{x^2}{2} \right) \pi_{b_0}(x) dx.$$

This implies finiteness of the first term on the right-hand side of the last equality in (12), and hence of  $\text{KL}(b_0, b)$  too.  $\square$

The next lemma is a restatement of Lemma 5.2 in van der Meulen and van Zanten (2013), cf. also the proof of formula (2.1) in Tang and Ghosal (2007).

**Lemma A.2.** *Suppose that a prior  $\Pi$  has property (11). If for a sequence  $C_n$  of measurable subsets of  $\mathcal{X}(K)$  there exists a constant  $c > 0$ , such that*

$$e^{nc} \int_{C_n} L_n(b) \Pi(db) \rightarrow 0, \quad P_{b_0}\text{-a.s.},$$

*then*

$$(13) \quad \Pi(C_n | X_0, \dots, X_{\Delta n}) \rightarrow 0, \quad P_{b_0}\text{-a.s.}$$

*as  $n \rightarrow \infty$ .*



*Proof.* The proof of Lemma 5.2 in van der Meulen and van Zanten (2013) remains applicable. The required version of the strong law of large numbers for ergodic sequences invoked in van der Meulen and van Zanten (2013) follows for instance from Theorems 3.5.8 and 3.5.7 in Stout (1974).  $\square$

The next lemma is a restatement of Lemma 5.3 in van der Meulen and van Zanten (2013), the proof of which employs some arguments from Tang and Ghosal (2007).

**Lemma A.3.** *Fix  $\varepsilon > 0$  such that  $\varepsilon \leq 2\nu(\mathbb{R})$ , take a fixed  $f \in \mathcal{C}_{bdd}(\mathbb{R})$  such that  $\|f\|_\infty \leq 1$ , and write*

$$B = \{b \in \mathcal{X}(K) : \|P_\Delta^b f - P_\Delta^{b_0} f\|_{1,\nu} > \varepsilon\}.$$

*Then there exist a compact set  $F \subset \mathbb{R}$ , an integer  $N > 0$  and bounded intervals  $I_1, \dots, I_N$  covering  $F$ , such that*

$$B \subset \left( \bigcup_{j=1}^N B_j^+ \right) \cup \left( \bigcup_{j=1}^N B_j^- \right),$$

*where*

$$B_j^+ = \left\{ b \in B : P_\Delta^b f(x) - P_\Delta^{b_0} f(x) > \frac{\varepsilon}{4\nu(F)}, \forall x \in I_j \right\},$$

$$B_j^- = \left\{ b \in B : P_\Delta^b f(x) - P_\Delta^{b_0} f(x) < -\frac{\varepsilon}{4\nu(F)}, \forall x \in I_j \right\}.$$

*Proof.* The proof of this lemma requires Lemma A.4 that corresponds to Lemma A.1 in van der Meulen and van Zanten (2013). The arguments from the proof of Lemma 5.3 in van der Meulen and van Zanten (2013) carry over.  $\square$

The next lemma is an adaptation of Lemma A.1 in van der Meulen and van Zanten (2013), but in its proof we need somewhat different arguments than those used in van der Meulen and van Zanten (2013).

**Lemma A.4.** *For every fixed  $f \in \mathcal{C}_{bdd}(\mathbb{R})$ , the family  $\{P_\Delta^b f : b \in \mathcal{X}(K)\}$  is locally uniformly equicontinuous.*

*Proof.* By definition we need to show that the family  $\{P_\Delta^b f : b \in \mathcal{X}(K)\}$  is uniformly equicontinuous whenever the argument  $x$  of  $P_\Delta^b f(x)$  is restricted to an arbitrary compact set  $F$ . Since the transition operators form a semigroup, it is enough to prove the latter claim for  $\Delta$  small enough, in particular for  $\Delta$  satisfying

$$(14) \quad K\Delta < \frac{1}{2}.$$

In fact, we have

$$|P_\Delta^b f(x) - P_\Delta^b f(y)| \leq P_{\Delta/2}^b |P_{\Delta/2}^b f(x) - P_{\Delta/2}^b f(y)|,$$

and if  $\{P_{\Delta/2}^b f : b \in \mathcal{X}(K)\}$  is uniformly equicontinuous when the argument  $x$  ranges in  $F$ , then it is immediately seen that so is  $\{P_\Delta^b f : b \in \mathcal{X}(K)\}$ , while if not, then we can reiterate the same argument, but now with  $\Delta/2$  and  $\Delta/4$  instead of  $\Delta$  and  $\Delta/2$  and so on, until (14) is met.

Fix a compact set  $F$  and let

$$l_u = \int_0^\Delta b(u + W_s) dW_s - \frac{1}{2} \int_0^\Delta b^2(u + W_s) ds, \quad L_u = e^{l_u}$$

for a standard Brownian motion  $W$ . Then as in van der Meulen and van Zanten (2013), it can be shown that

$$P_\Delta^b f(x) = \mathbb{E}[f(x + W_\Delta)L_x],$$

where the expectation is evaluated under the Wiener measure (the Girsanov theorem invoked in van der Meulen and van Zanten (2013) is applicable in our case thanks to the linear growth condition and Corollary 5.16 on p. 200 in Karatzas and Shreve (1988)). Also

$$\begin{aligned} |P_\Delta^b f(x) - P_\Delta^b f(y)| &\leq \mathbb{E}[|f(x + W_\Delta)| |L_x - L_y|] + \mathbb{E}[L_y |f(x + W_\Delta) - f(y + W_\Delta)|] \\ &:= S_1 + S_2, \end{aligned}$$

where  $x, y \in F$ . We will bound the two terms  $S_1$  and  $S_2$  separately.

By (14) there exists  $\tilde{q} > 1$ , such that

$$(15) \quad K\Delta < \frac{1}{2\sqrt{\tilde{q}}}.$$

Fix such  $\tilde{q}$  and let  $q$  be determined as that root of the equation

$$(16) \quad \tilde{q} = 2 \left( q^2 - \frac{q}{2} \right)$$

that is larger than 1. Next set  $r = q/(q-1)$ . Note that  $r > 1$  and that  $1/r + 1/q = 1$ .

To bound  $S_1$ , we apply an elementary inequality  $|e^a - e^b| \leq |a - b|e^a + e^b$  for  $a, b \in \mathbb{R}$  and Hölder's inequality with exponents  $r$  and  $q$  defined as above to obtain

$$\begin{aligned} S_1 &\leq \|f\|_\infty \mathbb{E}[|L_x - L_y|] \\ &\leq \|f\|_\infty \mathbb{E}[|l_x - l_y| |L_x + L_y|] \\ &\leq \|f\|_\infty \{\mathbb{E}[|l_x - l_y|^r]\}^{1/r} \{\mathbb{E}[|L_x + L_y|^q]\}^{1/q}. \end{aligned}$$

In order to bound  $S_1$ , we hence need to bound the last two factors on the right-hand side of the last inequality in the above display. We first treat the first of these two. Note that

$$\begin{aligned} (17) \quad l_x - l_y &= \int_0^\Delta (b(x + W_s) - b(y + W_s)) dW_s - \frac{1}{2} \int_0^\Delta (b^2(x + W_s) - b^2(y + W_s)) ds \\ &:= S_3 + S_4. \end{aligned}$$

The  $c_r$ -inequality yields that in order to bound  $\{\mathbb{E}[|l_x - l_y|^r]\}^{1/r}$ , it is enough to bound  $\mathbb{E}[|S_3|^r]$  and  $\mathbb{E}[|S_4|^r]$ . We bound the first of these two expectations as follows: by the Burkholder-Davis-Gundy inequality, see Theorem 3.28 on p. 166 in Karatzas and Shreve (1988),

$$\begin{aligned} \mathbb{E} \left[ \left| \int_0^\Delta (b(x + W_s) - b(y + W_s)) dW_s \right|^r \right] \\ \leq C_r \mathbb{E} \left[ \left| \int_0^\Delta (b(x + W_s) - b(y + W_s))^2 ds \right|^{r/2} \right], \end{aligned}$$

where  $C_r > 0$  is a universal constant independent of  $b, x$  and  $y$ . For a constant  $R > 0$  and the set  $F' = \{u + v : u \in F, v \in [-R, R]\}$  by the Cauchy-Schwarz inequality the expectation on the right-hand side of the above display can be handled as follows:

$$\begin{aligned}
& \mathbb{E} \left[ \left| \int_0^\Delta (b(x + W_s) - b(y + W_s))^2 ds \right|^{r/2} 1_{[\sup_{s \leq \Delta} |W_s| \leq R]} \right] \\
& + \mathbb{E} \left[ \left| \int_0^\Delta (b(x + W_s) - b(y + W_s))^2 ds \right|^{r/2} 1_{[\sup_{s \leq \Delta} |W_s| > R]} \right] \\
& \leq \Delta^{r/2} \sup_{\substack{u, v \in F' \\ |u-v| \leq |x-y|}} |b(u) - b(v)|^r \\
& + \left\{ \mathbb{E} \left[ \left| \int_0^\Delta (b(x + W_s) - b(y + W_s))^2 ds \right|^r \right] \right\}^{1/2} \left\{ P \left( \sup_{s \leq \Delta} |W_s| > R \right) \right\}^{1/2}.
\end{aligned}$$

Thanks to the fact that  $\mathcal{X}(K)$  is a locally uniformly equicontinuous family of functions, for a fixed  $R$  the first term on the right-hand side of the last inequality can be made arbitrarily small uniformly in  $b \in \mathcal{X}(K)$  by choosing  $\delta$  small enough and  $|x - y| \leq \delta$ . Also  $P(\sup_{s \leq \Delta} |W_s| > R)$  can be made arbitrarily small by choosing  $R$  large enough. Finally,

$$\begin{aligned}
& \mathbb{E} \left[ \left| \int_0^\Delta (b(x + W_s) - b(y + W_s))^2 ds \right|^r \right] \\
& \leq \Delta^{r/q} \mathbb{E} \left[ \int_0^\Delta |b(x + W_s) - b(y + W_s)|^{2r} ds \right].
\end{aligned}$$

The expectation on the right-hand side is bounded by a universal constant independent of particular  $x, y \in F$  and  $b \in \mathcal{X}(K)$ . This can be seen by a simple, but lengthy computation employing the Fubini theorem, the linear growth condition on  $b$ , the  $c_{2r}$ -inequality and the fact that  $W$  has moments of all orders. This completes bounding  $\mathbb{E}[|S_3|^r]$ , which hence can be made arbitrarily small uniformly in  $b \in \mathcal{X}(K)$ , once  $|x - y| \leq \delta$  for  $\delta > 0$  small enough. The term  $\mathbb{E}[|S_4|^r]$  can be bounded using similar computations by employing the inequality

$$|b^2(u) - b^2(v)| \leq K(2 + |u| + |v|)|b(u) - b(v)|$$

and the Cauchy-Schwarz inequality (twice), yielding together

$$\begin{aligned}
& \mathbb{E} \left[ \left| \int_0^\Delta (b^2(x + W_s) - b^2(y + W_s)) ds \right|^r \right] \\
& \leq \left\{ \mathbb{E} \left[ \left| \int_0^\Delta (b(x + W_s) - b(y + W_s))^2 ds \right|^r \right] \right\}^{1/2} \\
& \quad \times \left\{ \mathbb{E} \left[ \left| \int_0^\Delta (b(x + W_s) + b(y + W_s))^2 ds \right|^r \right] \right\}^{1/2}.
\end{aligned}$$

The first factor on the right-hand side can be made arbitrarily small uniformly in  $b \in \mathcal{X}(K)$  by taking  $\delta$  small (cf. above), while the second factor remains bounded uniformly in  $b \in \mathcal{X}(K)$  and  $x, y \in F$ . To complete bounding  $S_1$ , we need to bound the right-hand side of the inequality

$$\mathbb{E}[|L_x + L_y|^q] \leq c_q \mathbb{E}[L_x^q] + c_q \mathbb{E}[L_y^q].$$

Since obviously both terms on the right-hand side can be bounded in exactly the same manner, we will only give an argument for one of them. By the Cauchy-Schwarz inequality applied to two random variables

$$\begin{aligned} & \exp \left( \left( q^2 - \frac{q}{2} \right) \int_0^\Delta b^2(x + W_s) ds \right), \\ & \exp \left( \int_0^\Delta q b(x + W_s) dW_s - \int_0^\Delta q^2 b^2(x + W_s) ds \right), \end{aligned}$$

we have

$$(18) \quad \mathbb{E}[L_x^q] \leq \left\{ \mathbb{E} \left[ \exp \left( 2 \left( q^2 - \frac{q}{2} \right) \int_0^\Delta b^2(x + W_s) ds \right) \right] \right\}^{1/2}.$$

Here we used the fact that

$$(19) \quad \mathbb{E} \left[ \exp \left( \int_0^\Delta 2qb(x + W_s) dW_s - \frac{1}{2} \int_0^\Delta 4q^2 b^2(x + W_s) ds \right) \right] = 1,$$

since the process under the expectation sign is a martingale and has expectation equal to one (this is due to the linear growth condition and Corollary 5.16 on p. 200 in Karatzas and Shreve (1988)). Hence it remains to bound the right-hand side of (18), which we denote by  $S_5$ . By the linear growth condition we have

$$S_5^2 \leq \exp(2\tilde{q}K^2\Delta(1 + |x|)^2) \mathbb{E} \left[ \exp \left( 2\tilde{q}K^2 \int_0^\Delta W_s^2 ds \right) \right].$$

Showing finiteness of the expectation on the right-hand side is standard: by Doob's maximal inequality for submartingales (see Theorem 3.8 (iv) pp. 13–14 in Karatzas and Shreve (1988); that the exponential on the right-hand side of the first line of the displayed formula below is a submartingale follows from Problem 3.7 on p. 13 in Karatzas and Shreve (1988)),

$$\begin{aligned} \mathbb{E} \left[ \exp \left( 2\tilde{q}K^2 \int_0^\Delta W_s^2 ds \right) \right] & \leq \mathbb{E} \left[ \sup_{s \leq \Delta} \exp(2\tilde{q}K^2\Delta W_s^2) \right] \\ & \leq 4\mathbb{E} [\exp(2\tilde{q}K^2\Delta W_\Delta^2)] < \infty. \end{aligned}$$

Here in the last inequality we used (15).

A conclusion that follows from the above bounds is that  $S_1$  can be made arbitrarily small as soon as  $|x - y| \leq \delta$  for small enough  $\delta$ . The bound on  $S_1$  will be true uniformly in  $b \in \mathcal{X}(K)$ .

In order to bound  $S_2$ , we again use Hölder's inequality to get

$$S_2 \leq \{\mathbb{E}[L_y^q]\}^{1/q} \{\mathbb{E}[|f(x + W_\Delta) - f(y + W_\Delta)|^r]\}^{1/r}.$$

The first factor on the right-hand side can be bounded as above. The second factor can be made arbitrarily small as soon as  $|x - y| \leq \delta$  for small enough  $\delta$ . Indeed, for a constant  $R > 0$  write

$$\begin{aligned} \mathbb{E}[|f(x + W_\Delta) - f(y + W_\Delta)|^r] & = \mathbb{E}[|f(x + W_\Delta) - f(y + W_\Delta)|^r 1_{|W_\Delta| > R}] \\ & \quad + \mathbb{E}[|f(x + W_\Delta) - f(y + W_\Delta)|^r 1_{|W_\Delta| \leq R}] \\ & \leq (2\|f\|_\infty)^r P(|W_\Delta| > R) \\ & \quad + \mathbb{E}[|f(x + W_\Delta) - f(y + W_\Delta)|^r 1_{|W_\Delta| \leq R}]. \end{aligned}$$

It is obvious that the first term on the right-hand side of the last inequality can be made arbitrarily small by selecting  $R$  large enough. However, so can be the second one upon fixing  $R$  by taking  $|x - y| \leq \delta$  for small enough  $\delta > 0$ , since the function  $f$  is uniformly continuous on compacts. This completes the proof.  $\square$

## APPENDIX B.

**Lemma B.1.** *Let  $b, \tilde{b} \in \tilde{\mathcal{X}}(K_1, K_2)$ . Fix  $t > 0$ . If  $b \neq \tilde{b}$ , then  $P_t^b \neq P_t^{\tilde{b}}$ .*

*Proof.* The proof is similar to the proof of Lemma 3.1 in van der Meulen and van Zanten (2013). By continuity of  $b$  and  $\tilde{b}$  we have that if  $b \neq \tilde{b}$ , this in fact holds on a set of positive Lebesgue measure. Then also  $V_b \neq V_{\tilde{b}}$  on a set of positive Lebesgue measure and therefore  $\pi_b \neq \pi_{\tilde{b}}$  on a set of positive Lebesgue measure, for instance some open ball in  $\mathbb{R}^d$ . Now assume that  $P_t^b = P_t^{\tilde{b}}$ . Then for any bounded measurable function  $f$  and any positive integer  $m$ , by the semigroup property of  $P_t^b$  we have that

$$\mathbb{E}_x^b[f(X_{mt})] = (P_t^b)^m f(x) = (P_t^{\tilde{b}})^m f(x) = \mathbb{E}_x^{\tilde{b}}[f(X_{mt})].$$

Letting  $m \rightarrow \infty$ , the above display and ergodicity give that

$$\int_{\mathbb{R}^d} f(y) \pi_b(y) dy = \int_{\mathbb{R}^d} f(y) \pi_{\tilde{b}}(y) dy.$$

Hence  $\pi_b = \pi_{\tilde{b}}$  a.e., and in fact by continuity  $\pi_b = \pi_{\tilde{b}}$  everywhere. This is a contradiction and thus  $b \neq \tilde{b}$  implies  $P_t^b \neq P_t^{\tilde{b}}$ .  $\square$

**Lemma B.2.** *Fix  $\varepsilon > 0$  such that  $\varepsilon \leq 2\nu(\mathbb{R})$ , take a fixed  $f \in \mathcal{C}_{bdd}(\mathbb{R}^d)$  such that  $\|f\|_\infty \leq 1$ , and write*

$$B = \{b \in \mathcal{X} : \|P_\Delta^b f - P_\Delta^{b_0} f\|_{1,\nu} > \varepsilon\}.$$

*Then there exist a compact set  $F \subset \mathbb{R}^d$ , an integer  $N > 0$  and cubes  $I_1, \dots, I_N$  covering  $F$ , such that*

$$B \subset \left( \bigcup_{j=1}^N B_j^+ \right) \cup \left( \bigcup_{j=1}^N B_j^- \right),$$

*where*

$$B_j^+ = \left\{ b \in B : P_\Delta^b f(x) - P_\Delta^{b_0} f(x) > \frac{\varepsilon}{4\nu(F)}, \forall x \in I_j \right\},$$

$$B_j^- = \left\{ b \in B : P_\Delta^b f(x) - P_\Delta^{b_0} f(x) < -\frac{\varepsilon}{4\nu(F)}, \forall x \in I_j \right\}.$$

*Proof.* The proof of Lemma 5.3 in van der Meulen and van Zanten (2013) carries over, provided one redefines the intervals  $I_j$  of length  $\delta/2 > 0$  from that proof to be cubes with sides of length  $\delta/2$ , and uses instead of Lemma A.1 from van der Meulen and van Zanten (2013) Lemma B.3 below.  $\square$

**Lemma B.3.** *For a fixed  $f \in \mathcal{C}_{bdd}(\mathbb{R}^d)$  and  $t > 0$ , the family  $\{P_t^b f : b \in \mathcal{X}(K_1, K_2)\}$  is a locally uniformly equicontinuous family of functions.*

*Proof.* Let

$$l_u = \sum_{i=1}^d \int_0^\Delta b_i(u + W_s) dW_{i,s} - \frac{1}{2} \sum_{i=1}^d \int_0^\Delta b_i^2(u + W_s) ds, \quad L_u = e^{l_u}$$

for a standard  $d$ -dimensional Brownian motion  $W = (W_1, \dots, W_d)$ . Then, employing the Girsanov theorem, as in the proof of Lemma A.1 in van der Meulen and van Zanten (2013), see also the proof of Lemma A.4, it can be shown that

$$P_\Delta^b f(x) = \mathbb{E}[f(x + W_\Delta) L_x],$$

where the expectation is evaluated under the Wiener measure. From this point on the proof is a generalisation of the arguments in the proof of Lemma A.4 from Appendix A to the multidimensional setting. In particular, as in that proof, it is enough to prove the lemma for  $\Delta$  such that

$$\Delta K_1 < \frac{1}{2\sqrt{d}}.$$

In order to prove the lemma, we need to show that the family of functions  $\{P_t^b f : b \in \mathcal{X}(K_1, K_2)\}$  is uniformly equicontinuous whenever the argument  $x$  of  $P_t^b f(x)$  is restricted to an arbitray compact set  $F$ . Fix a compact set  $F \subset \mathbb{R}^d$ . Throughout this proof we assume  $x, y \in F$ . We have

$$\begin{aligned} |P_\Delta^b f(x) - P_\Delta^b f(y)| &\leq \mathbb{E}[|f(x + W_\Delta) - f(y + W_\Delta)| L_x] + \mathbb{E}[L_y |f(x + W_\Delta) - f(y + W_\Delta)|] \\ &:= S_1 + S_2. \end{aligned}$$

We will bound the two terms  $S_1$  and  $S_2$  separately. There exists  $\tilde{q} > 1$ , such that

$$(20) \quad K_1 \Delta < \frac{1}{2\sqrt{d\tilde{q}}}.$$

Fix such  $\tilde{q}$  and let  $q$  be determined as that root of the equation

$$(21) \quad \tilde{q} = 2 \left( q^2 - \frac{q}{2} \right)$$

that is larger than 1. Next set  $r = q/(q-1)$ . We first bound  $S_1$ . As in the proof of Lemma A.4 from Appendix A, we have that

$$S_1 \leq \|f\|_\infty \{\mathbb{E}[|l_x - l_y|^r]\}^{1/r} \{\mathbb{E}[|L_x + L_y|^q]\}^{1/q}.$$

The  $c_r$ -inequality gives that in order to bound  $\{\mathbb{E}[|l_x - l_y|^r]\}^{1/r}$ , it is enough to bound the terms

$$\begin{aligned} &\mathbb{E} \left[ \left| \int_0^\Delta (b_i(x + W_s) - b_i(y + W_s)) dW_{i,s} \right|^r \right], \\ &\mathbb{E} \left[ \left| \int_0^\Delta (b_i^2(x + W_s) - b_i^2(y + W_s)) ds \right|^r \right] \end{aligned}$$

for  $i = 1, \dots, d$ . Since the arguments are the same for any  $i$ , we henceforth fix a particular  $i$ . As in Lemma A.4, the Burkholder-Davis-Gundy inequality gives

$$\mathbb{E} \left[ \left| \int_0^\Delta (b_i(x + W_s) - b_i(y + W_s)) dW_{i,s} \right|^r \right]$$

$$\leq C_r \mathbb{E} \left[ \left| \int_0^\Delta (b_i(x + W_s) - b_i(y + W_s))^2 ds \right|^{r/2} \right],$$

where  $C_r > 0$  is a universal constant independent of  $b$ . For a constant  $R > 0$  and the set  $F' = \{u + v : u \in F, \|v\| \leq R\}$  by the Cauchy-Schwarz inequality the expectation on the right-hand side of the above display can be bounded as follows:

$$\begin{aligned} & \mathbb{E} \left[ \left| \int_0^\Delta (b_i(x + W_s) - b_i(y + W_s))^2 ds \right|^{r/2} 1_{[\sup_{s \leq \Delta} \|W_s\| \leq R]} \right] \\ & + \mathbb{E} \left[ \left| \int_0^\Delta (b_i(x + W_s) - b_i(y + W_s))^2 ds \right|^{r/2} 1_{[\sup_{s \leq \Delta} \|W_s\| > R]} \right] \\ & \leq \Delta^{r/2} \sup_{\substack{u, v \in F' \\ \|u - v\| \leq \|x - y\|}} |b_i(u) - b_i(v)|^r \\ & + \left\{ \mathbb{E} \left[ \left| \int_0^\Delta (b_i(x + W_s) - b_i(y + W_s))^2 ds \right|^r \right] \right\}^{1/2} \left\{ P \left( \sup_{s \leq \Delta} \|W_s\| > R \right) \right\}^{1/2}. \end{aligned}$$

Since  $b$  has partial derivatives bounded in absolute value by  $K_2$ , the first term on the right-hand side of the above display can be made arbitrarily small by choosing  $\delta$  small enough and  $\|x - y\| \leq \delta$ . Furthermore, the term

$$\left\{ P \left( \sup_{s \leq \Delta} \|W_s\| > R \right) \right\}^{1/2}$$

can be made arbitrarily small by choosing  $R$  large enough. A lengthy, but easy computation shows that the term

$$\left\{ \mathbb{E} \left[ \left| \int_0^\Delta (b_i(x + W_s) - b_i(y + W_s))^2 ds \right|^r \right] \right\}^{1/2}$$

is bounded by a constant independent of  $b$ ; cf. the arguments in the proof of Lemma A.4 from Appendix A. Consequently, the term

$$\mathbb{E} \left[ \left| \int_0^\Delta (b_i(x + W_s) - b_i(y + W_s)) dW_{i,s} \right|^r \right]$$

can be made arbitrarily small, once  $\delta$  is chosen small enough and  $\|x - y\| \leq \delta$ . The term

$$\mathbb{E} \left[ \left| \int_0^\Delta (b_i^2(x + W_s) - b_i^2(y + W_s)) ds \right|^r \right]$$

can be shown to be bounded uniformly in  $b \in \mathcal{X}(K_1, K_2)$  by employing similar techniques; cf. the proof of Lemma A.4 from Appendix A. Next we need to bound the right-hand side of the inequality

$$\mathbb{E} [|L_x + L_y|^q] \leq c_q \mathbb{E} [L_x^q] + c_q \mathbb{E} [L_y^q].$$

Since obviously both terms on the right-hand side can be bounded in exactly the same manner, we will only give an argument for the first one of them. By the

Cauchy-Schwarz inequality applied to the random variables

$$\exp \left( \left( q^2 - \frac{q}{2} \right) \sum_{i=1}^d \int_0^\Delta b_i^2(x + W_s) ds \right),$$

$$\exp \left( \sum_{i=1}^d \int_0^\Delta q b_i(x + W_s) dW_{i,s} - \sum_{i=1}^d \int_0^\Delta q^2 b_i^2(x + W_s) ds \right),$$

as in the proof of Lemma A.4 from Appendix A we have

$$(22) \quad \mathbb{E}[L_x^q] \leq \left\{ \mathbb{E} \left[ \exp \left( 2 \left( q^2 - \frac{q}{2} \right) \sum_{i=1}^d \int_0^\Delta b_i^2(x + W_s) ds \right) \right] \right\}^{1/2}.$$

Hence it remains to bound the right-hand side of the above display, which we denote by  $S_5$ . By the linear growth condition we have

$$S_5^2 \leq \exp(2d\tilde{q}K_1^2\Delta(1 + \|x\|^2)) \mathbb{E} \left[ \exp \left( 2d\tilde{q}K_1^2 \int_0^\Delta \|W_s\|^2 ds \right) \right].$$

By Doob's maximal inequality for submartingales and independence of scalar Brownian motions  $W_i$ 's,

$$\mathbb{E} \left[ \exp \left( 2d\tilde{q}K_1^2 \int_0^\Delta \|W_s\|^2 ds \right) \right] \leq 4 \prod_{i=1}^d \mathbb{E} [\exp(2d\tilde{q}K_1^2\Delta W_{i,\Delta}^2)] < \infty.$$

Here in the last inequality we used (20). The conclusion is that the term  $S_1$  can be made arbitrarily small by taking  $\delta$  small and  $\|x - y\| \leq \delta$ . The proof is now completed as in the case of Lemma A.4 from Appendix A: by Hölder's inequality

$$S_2 \leq \{\mathbb{E}[L_y^q]\}^{1/q} \{\mathbb{E}[|f(x + W_\Delta) - f(y + W_\Delta)|^r]\}^{1/r}.$$

The first factor on the right-hand side can be bounded as above uniformly in  $b \in \mathcal{X}(K_1, K_2)$ . The second factor can be made arbitrarily small as soon as  $\|x - y\| \leq \delta$  for small enough  $\delta$ : for a constant  $R > 0$ ,

$$\begin{aligned} \mathbb{E}[|f(x + W_\Delta) - f(y + W_\Delta)|^r] &= \mathbb{E}[|f(x + W_\Delta) - f(y + W_\Delta)|^r 1_{[\|W_\Delta\| > R]}] \\ &\quad + \mathbb{E}[|f(x + W_\Delta) - f(y + W_\Delta)|^r 1_{[\|W_\Delta\| \leq R]}] \\ &\leq (2\|f\|_\infty)^r P(\|W_\Delta\| > R) \\ &\quad + \mathbb{E}[|f(x + W_\Delta) - f(y + W_\Delta)|^r 1_{[\|W_\Delta\| \leq R]}]. \end{aligned}$$

The first term on the right-hand side of the last inequality can be made arbitrarily small by selecting  $R$  large enough. Upon fixing  $R$ , so can be the second one by taking  $\|x - y\| \leq \delta$  for small enough  $\delta > 0$ . Combination of all the above intermediate results entails the statement of the lemma.  $\square$

**Lemma B.4.** *Let*

$$\text{KL}(b_0, b) = \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \pi_{b_0}(x) p_{b_0}(\Delta, x, y) \log \frac{p_{b_0}(\Delta, x, y)}{p_b(\Delta, x, y)} dx dy.$$

*and assume that the weak solution to (1) is initialised at  $\mu_b$ . Then for the prior  $\Pi$  satisfying property (6), the inequality*

$$(23) \quad \Pi(b \in \mathcal{X}(K_1, K_2) : \text{KL}(b_0, b) < \varepsilon) > 0, \quad \forall \varepsilon > 0$$

*holds.*



*Proof.* The proof is an obvious modification of the proof of Lemma 5.1 in van der Meulen and van Zanten (2013); as in the proof of Lemma A.1 in Appendix A, we need to verify additionally that the Kullback-Leibler divergence  $K(\mu_b, \mu_{\tilde{b}})$  is finite for any  $b, \tilde{b} \in \mathcal{X}(K_1, K_2)$ . This, however, follows from Proposition 1.1 in Gobet (2002).  $\square$

**Lemma B.5.** *Suppose that the prior  $\Pi$  on  $\mathcal{X}(K_1, K_2)$  has the property (23) and assume that the weak solution to (1) is initialised at  $\mu_b$ . If for a sequence  $C_n$  of measurable subsets of  $\mathcal{X}(K_1, K_2)$  there exists a constant  $c > 0$ , such that*

$$e^{nc} \int_{C_n} L_n(b) \Pi(db) \rightarrow 0, \quad P_{b_0}\text{-a.s.},$$

*then*

$$\Pi(C_n | X_0, \dots, X_{\Delta n}) \rightarrow 0, \quad P_{b_0}\text{-a.s.}$$

*as  $n \rightarrow \infty$ .*

*Proof.* The proof is an easy generalisation of the proof of Lemma 5.2 in van der Meulen and van Zanten (2013).  $\square$

#### REFERENCES

- A. Beskos, O. Papaspiliopoulos, G. Roberts and P. Fearnhead. Exact and computationally efficient likelihood-based estimation for discretely observed diffusion processes. With discussions and a reply by the authors. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 68:333–382, 2006.
- L. Birgé. Robust testing for independent nonidentically distributed variables and Markov chains. *Specifying Statistical Models (Louvain-la-Neuve, 1981)*, pp. 134–162, Lecture Notes in Statist., 16, Springer, New York, 1983.
- F. Comte, V. Genon-Catalot and Y. Rozenholc. Penalized nonparametric mean square estimation of the coefficients of diffusion processes. *Bernoulli*, 13:514–543, 2007.
- D. Dacunha-Castelle and D. Florens-Zmirou. Estimation of the coefficients of a diffusion from discrete observations. *Stochastics*, 19:263–284, 1986.
- A. Dalalyan and M. Reiß. Asymptotic statistical equivalence for ergodic diffusions: the multidimensional case. *Probab. Theory Related Fields*, 137:25–47, 2007.
- P. Diaconis and D. Freedman. On the consistency of Bayes estimates. With a discussion and a rejoinder by the authors. *Ann. Statist.*, 14:1–67, 1986.
- R.M. Dudley. *Real Analysis and Probability*. Revised reprint of the 1989 original. Cambridge Studies in Advanced Mathematics, 74. Cambridge University Press, Cambridge, 2002.
- B. Eraker. MCMC analysis of diffusion models with application to finance. *J. Bus. Econom. Statist.*, 19:177–191, 2001.
- D. Florens-Zmirou. Approximate discrete-time schemes for statistics of diffusion processes. *Statistics*, 20:547–557, 1989.
- S. Ghosal, J.K. Ghosh and R.V. Ramamoorthi. Non-informative priors via sieves and packing numbers. *Advances in Statistical Decision Theory and Applications*, 119–132, Stat. Ind. Technol., Birkhäuser Boston, Boston, MA, 1997.
- S. Ghosal and Y. Tang. Bayesian consistency for Markov processes. *Sankhyā*, 68:227–239, 2006.

- S. Ghosal and A.W. van der Vaart. Convergence rates of posterior distributions for non-i.i.d. observations. *Ann. Statist.*, 35:192–223, 2007.
- E. Gobet. LAN property for ergodic diffusions with discrete observations. *Ann. Inst. H. Poincaré Probab. Statist.*, 38:711–737, 2002.
- E. Gobet, M. Hoffmann and M. Reiß. Nonparametric estimation of scalar diffusions based on low frequency data. *Ann. Statist.*, 32:2223–2253, 2004.
- J. Jacod. Non-parametric kernel estimation of the coefficient of a diffusion. *Scand. J. Statist.*, 27:83–96, 2000.
- I. Karatzas and S.E. Shreve. *Brownian Motion and Stochastic Calculus*. Graduate Texts in Mathematics, 113. Springer-Verlag, New York, 1988.
- F. van der Meulen, M. Schauer, H. van Zanten. Reversible jump MCMC for non-parametric drift estimation for diffusion processes. arXiv:1206.4910 [stat.CO], 2012.
- F.H. van der Meulen, A.W. van der Vaart and J.H. van Zanten. Convergence rates of posterior distributions for Brownian semimartingale models. *Bernoulli*, 12:863–888, 2006.
- F. van der Meulen and H. van Zanten. Consistent nonparametric Bayesian estimation for discretely observed scalar diffusions. *Bernoulli*, 19:44–63, 2013.
- L. Panzar and H. van Zanten. Nonparametric Bayesian inference for ergodic diffusions. *J. Statist. Plann. Inference*, 139:4193–4199, 2009.
- O. Papaspiliopoulos, Y. Pokern, G.O. Roberts and A.M. Stuart. Nonparametric estimation of diffusions: a differential equations approach. *Biometrika*, 99:511–531, 2012.
- Y. Pokern, A.M. Stuart and J.H. van Zanten. Posterior consistency via precision operators for nonparametric drift estimation in SDEs. *Stochastic Process. Appl.*, 123:603–628, 2013.
- D. Pollard. *A User’s Guide to Measure Theoretic Probability*. Cambridge Series in Statistical and Probabilistic Mathematics, 8. Cambridge University Press, Cambridge, 2002.
- L.C.G. Rogers and D. Williams. *Diffusions, Markov processes, and Martingales. Vol. 2. Itô calculus*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York, 1987.
- E. Schmisser. Penalized nonparametric drift estimation for a multidimensional diffusion process. *Statistics*, 47:61–84, 2013.
- A.V. Skorokhod. *Asimptoticheskie metody teorii stokhasticheskikh differentsial’nykh uravnenii (Russian) [Asymptotic Methods in the Theory of Stochastic Differential Equations]*. Naukova Dumka, Kiev, 1987.
- W.F. Stout. *Almost Sure Convergence*. Probability and Mathematical Statistics, Vol. 24. Academic Press [A subsidiary of Harcourt Brace Jovanovich, Publishers], New York-London, 1974.
- Y. Tang and S. Ghosal. Posterior consistency of Dirichlet mixtures for estimating a transition density. *J. Statist. Plann. Inference*, 137:1711–1726, 2007.
- S. Walker. On sufficient conditions for Bayesian consistency. *Biometrika*, 90:482–488, 2003.
- S. Walker. New approaches to Bayesian consistency. *Ann. Statist.*, 32:2028–2043, 2004.

- L. Wasserman. Asymptotic properties of nonparametric Bayesian procedures. *Practical Nonparametric and Semiparametric Bayesian Statistics*, 293–304, Lecture Notes in Statist., 133, Springer, New York, 1998.
- H. van Zanten. Nonparametric Bayesian methods for one-dimensional diffusion models: an overview of recent developments. arXiv:1209.6433 [**stat.ME**], 2012.

MATHEMATICAL INSTITUTE, LEIDEN UNIVERSITY, P.O. Box 9512, 2300 RA LEIDEN, THE NETHERLANDS

*E-mail address:* `shota.gugushvili@math.leidenuniv.nl`

KORTEWEG-DE VRIES INSTITUTE FOR MATHEMATICS, UNIVERSITEIT VAN AMSTERDAM, PO Box 94248, 1090 GE AMSTERDAM, THE NETHERLANDS

*E-mail address:* `spreij@uva.nl`